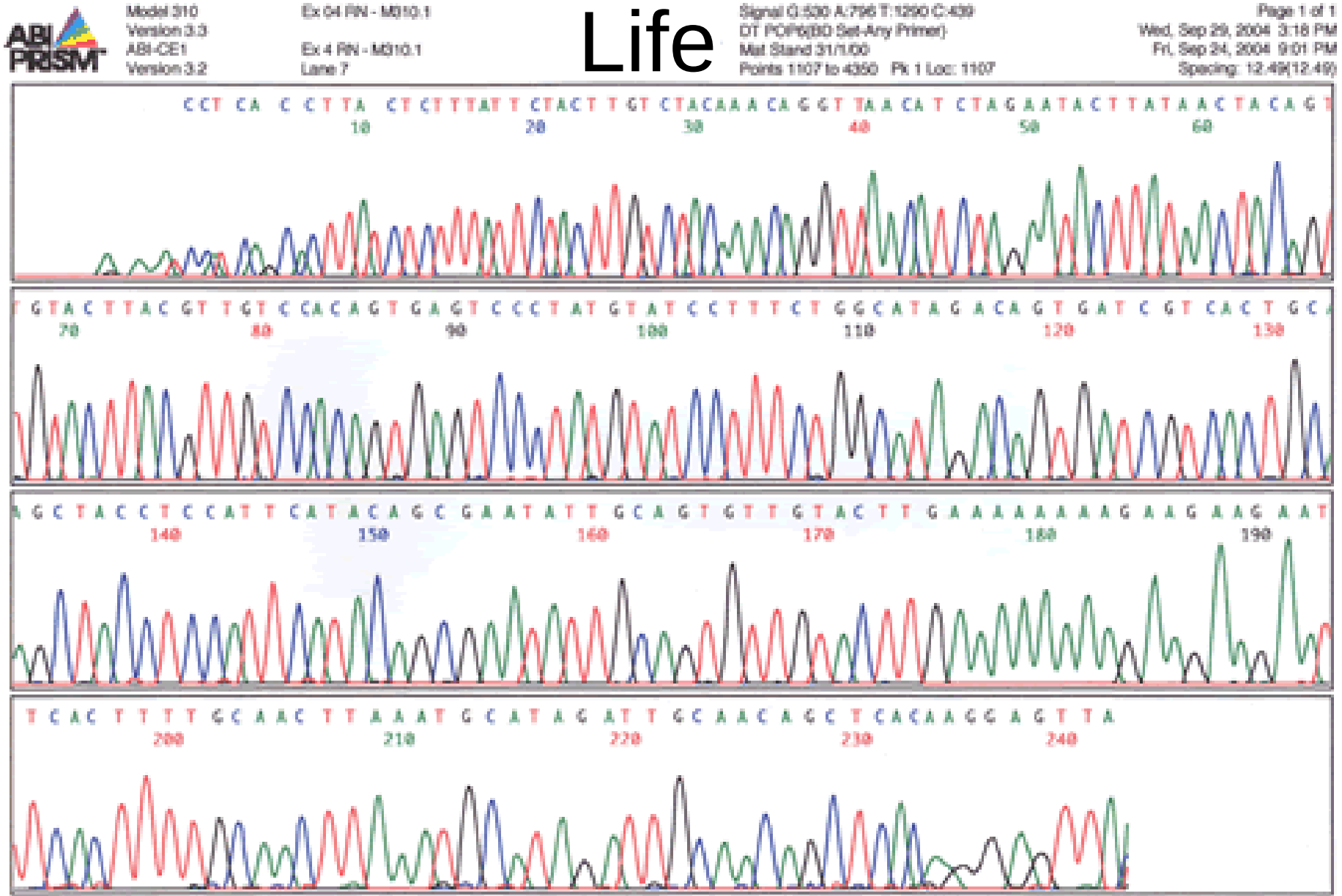
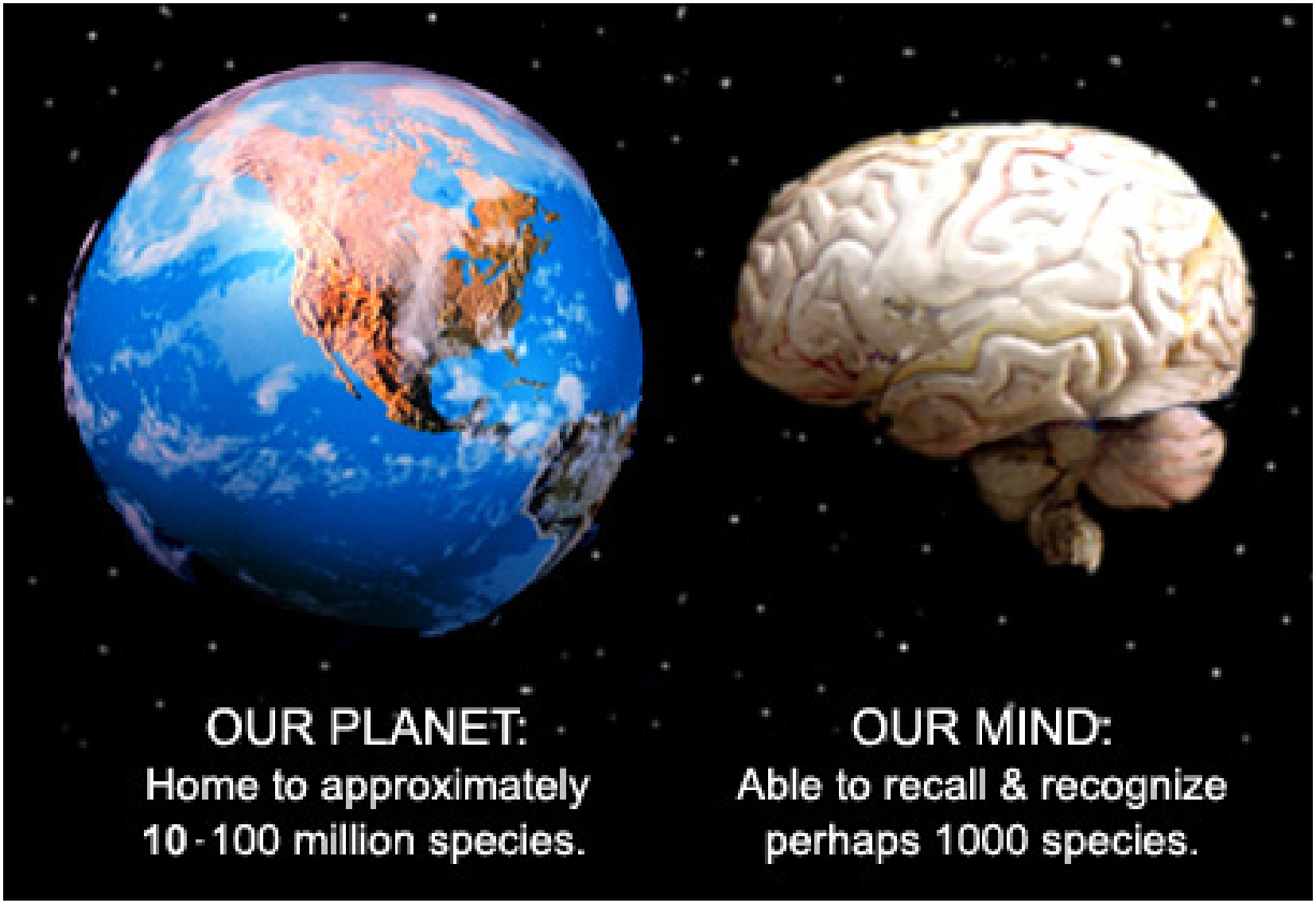
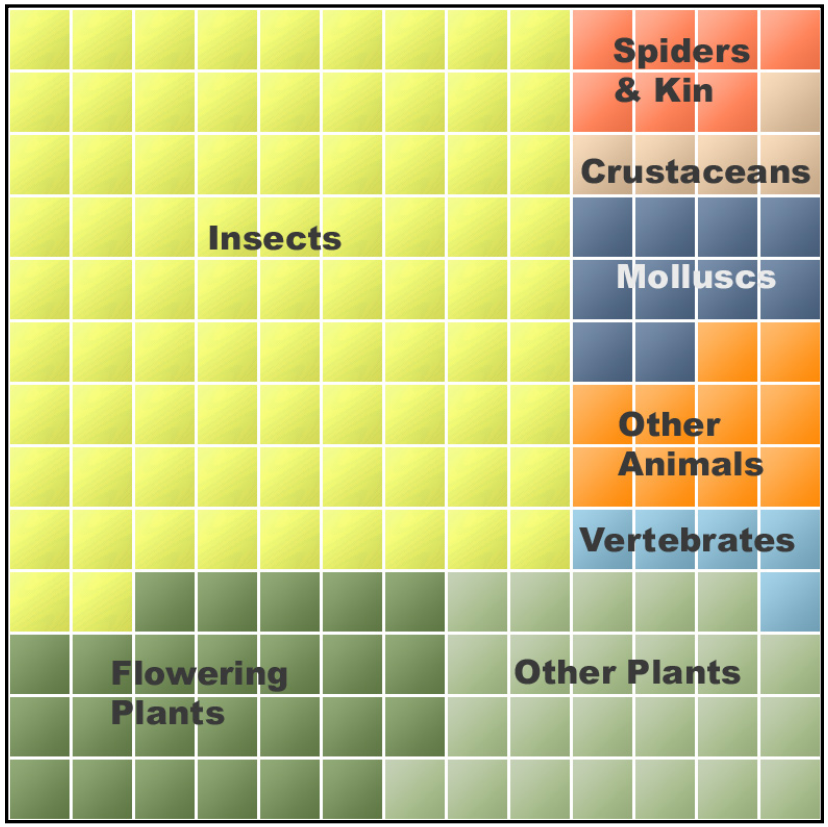
DNA Barcodes and Identifying Life



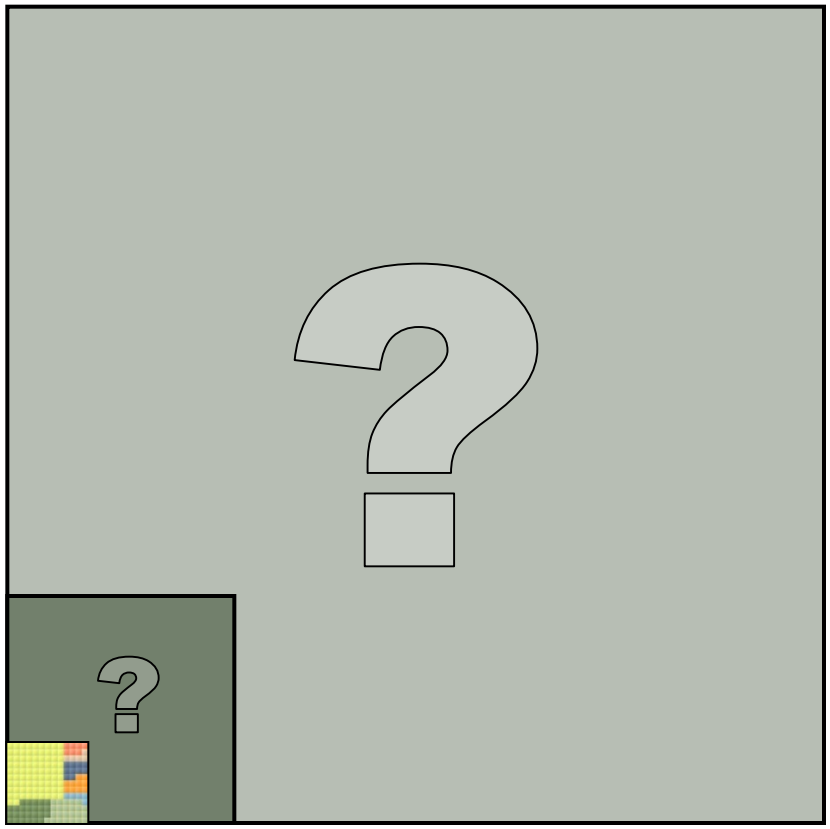
Identifying Life



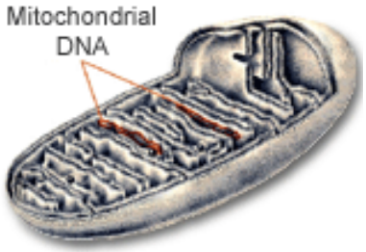
Known biodiversity: 1.7 million species of plants and animals



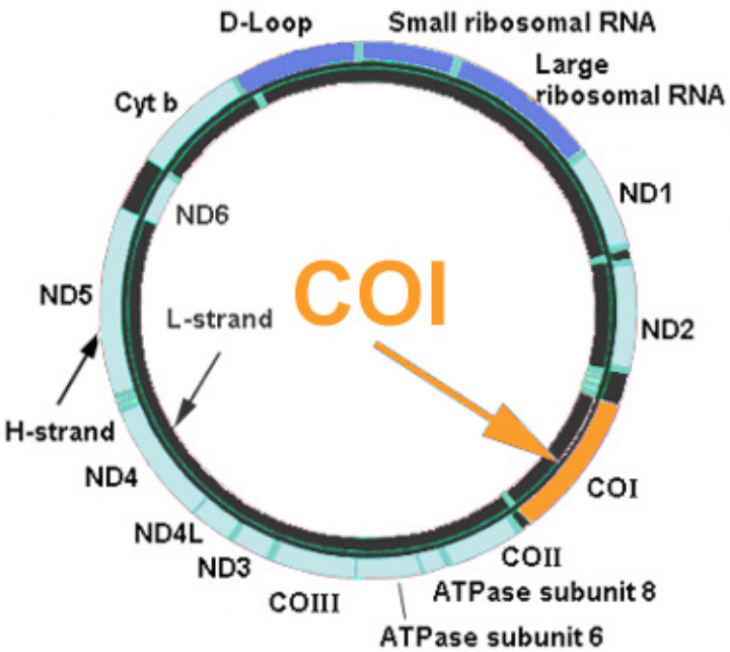
Estimated biodiversity: 10 million to 100 million species

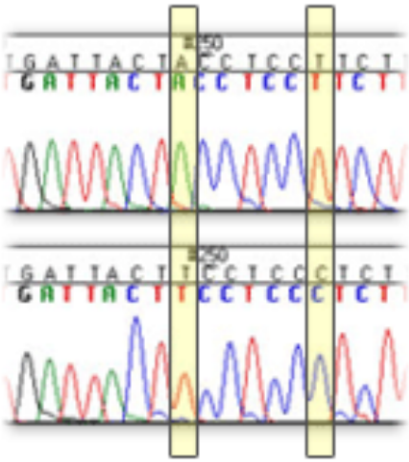


What and where is the COI gene?



Mitochondria are cellular organelles that contain their own stands of DNA





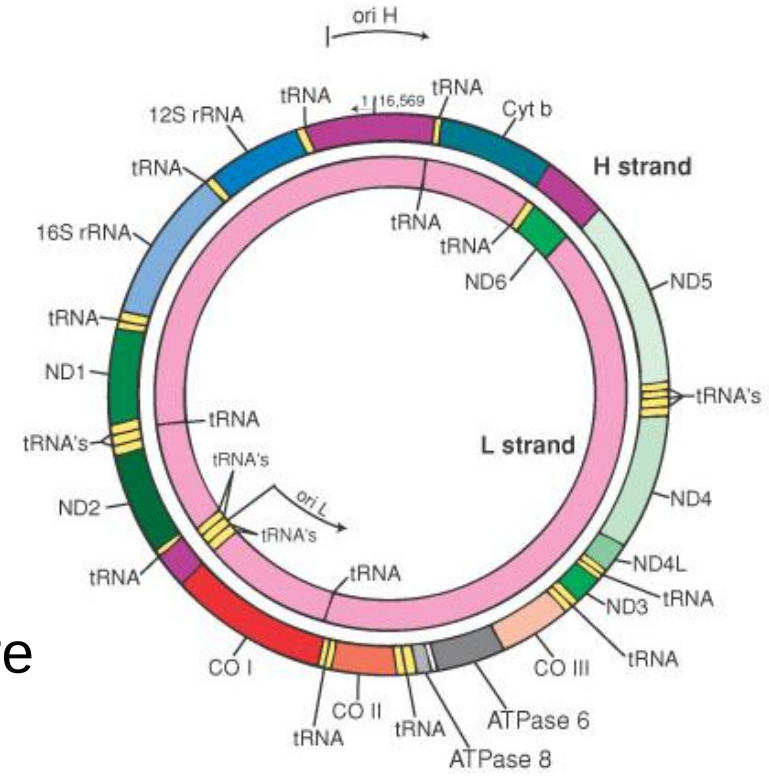
COI electropherograms provide a means of verifying key differences between species

COI is a mitochondrial protein coding gene necessary for oxidative metabolism.

Higher rates of 3rd position substitutions than amino acid composition changes; assumed modest rate of evolution at 2% divergence per million years.

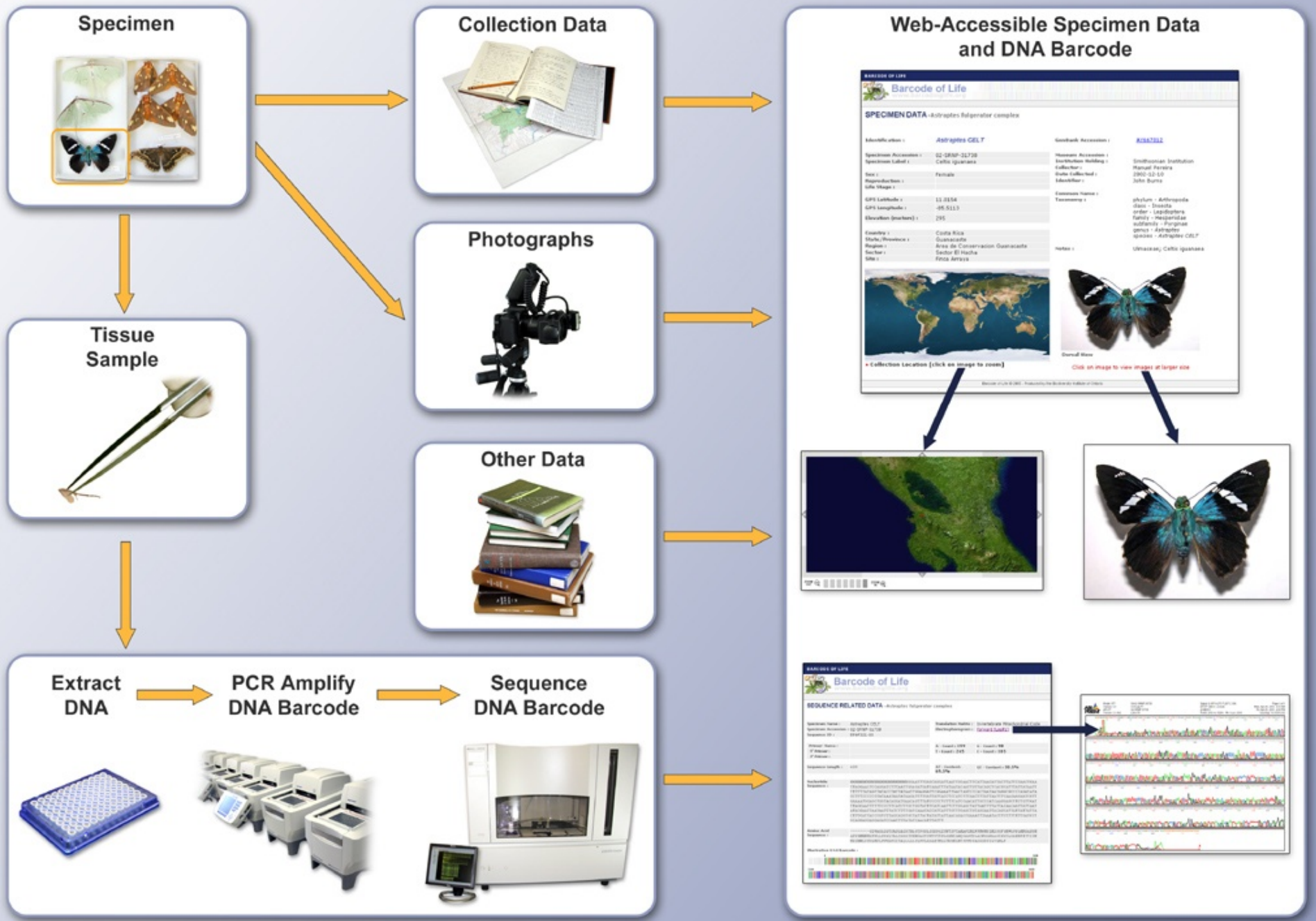
DNA barcoding is based on the use of a short, standard region that enables cost-effective species identification

* Cytochrome c oxidase I (COI) has become standard for animals
* “Folmer region" at the 5' end of gene
* closely related species often differ by several percent
* smaller N_e means gene coalesces (sorts) more quickly
* marker for both within species structure and among species structure

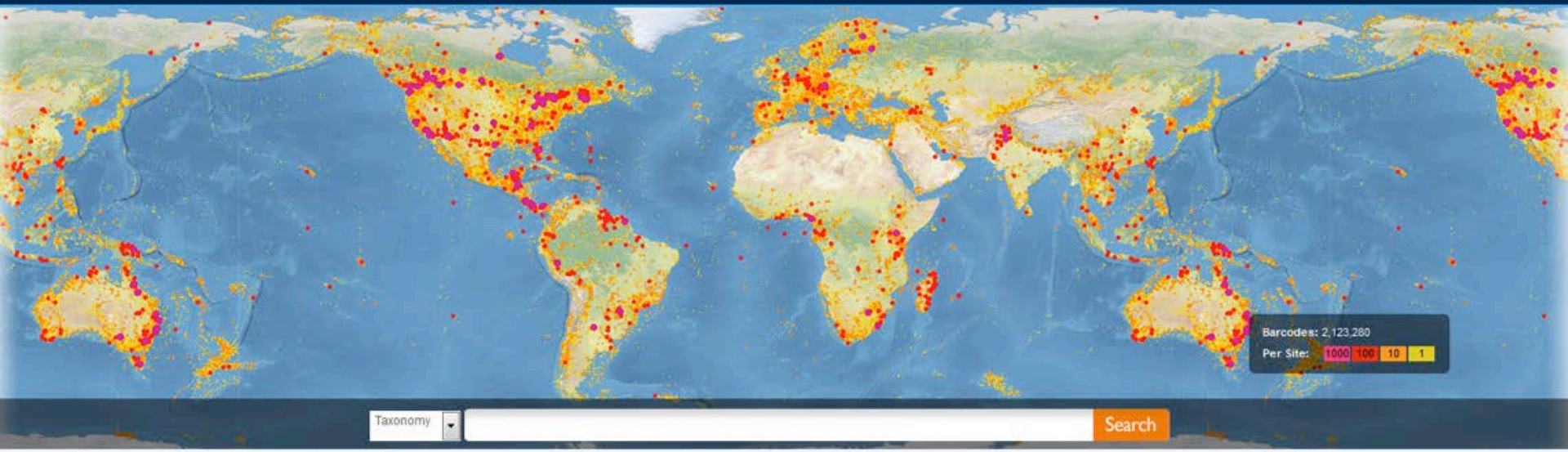


Horizontal Genomics – one gene for all species

DNA Barcoding: The Analytical Chain



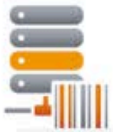
BOLDSYSTEMS Databases | Taxonomy | Identification | Workbench | Resources



The Barcode of Life Database is designed to support the generation and application of DNA Barcode data. The platform consists of three main modules, a data portal, a database of barcode clusters, and data collection workbench.

Public Data Portal

A data retrieval interface that allows for searching over 1.5M public records in BOLD using multiple search criteria including, but not limited to, geography, taxonomy, and depository.



BIN Database

A searchable database of Barcode Index Numbers (BINs), sequence clusters that closely approximate species.

Workbench

An integrated data collection and analysis environment that securely supports the assembly and validation of DNA barcodes and ancillary sequences.

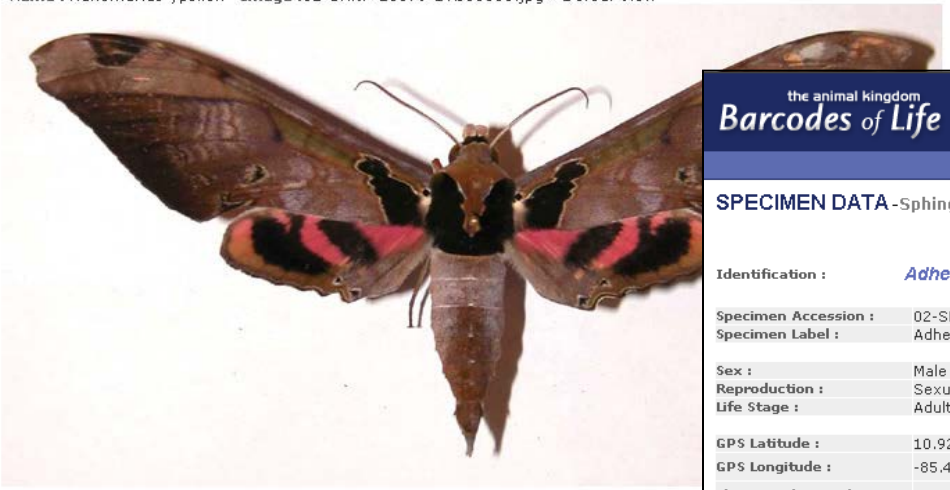
International Barcode of Life
Ontario Ministry of Economic Development and Innovation
Genome Canada
Ontario Genomics Institute
NSERC CRSNG

BOLD – Specimen Data

the animal kingdom
Barcodes of Life

:: Print
:: Help

Name : Adhemarius ypsilon Image :02-SRNP-18079-DHJ35550.jpg - Dorsal View



Name : Adhemarius ypsilon Image :02-SRNP-18079-DHJ35551.jpg - Ventral View

the animal kingdom
Barcodes of Life

:: Print
:: Help

SPECIMEN DATA - Sphingidae of Costa Rica

Edit

Identification : *Adhemarius ypsilon*

Specimen Accession :	02-SRNP-9473
Specimen Label :	Adhemarius ypsilon
Sex :	Male
Reproduction :	Sexual
Life Stage :	Adult
GPS Latitude :	10.9255
GPS Longitude :	-85.4716
Elevation (meters) :	
Country :	Costa Rica
State/Province :	Guanacaste
Region :	
Sector :	
Site :	

Museum Accession :	02-SRNP-9473
Institution Holding :	Smithsonian Institution
Collector :	Unknown
Date Collected :	2002-07-21
Identifier :	Dan Janzen
Common Name :	
Taxonomy :	phylum - Arthropoda class - Insecta order - Lepidoptera family - Sphingidae genus - *Adhemarius* species - *Adhemarius ypsilon*
Notes :	43

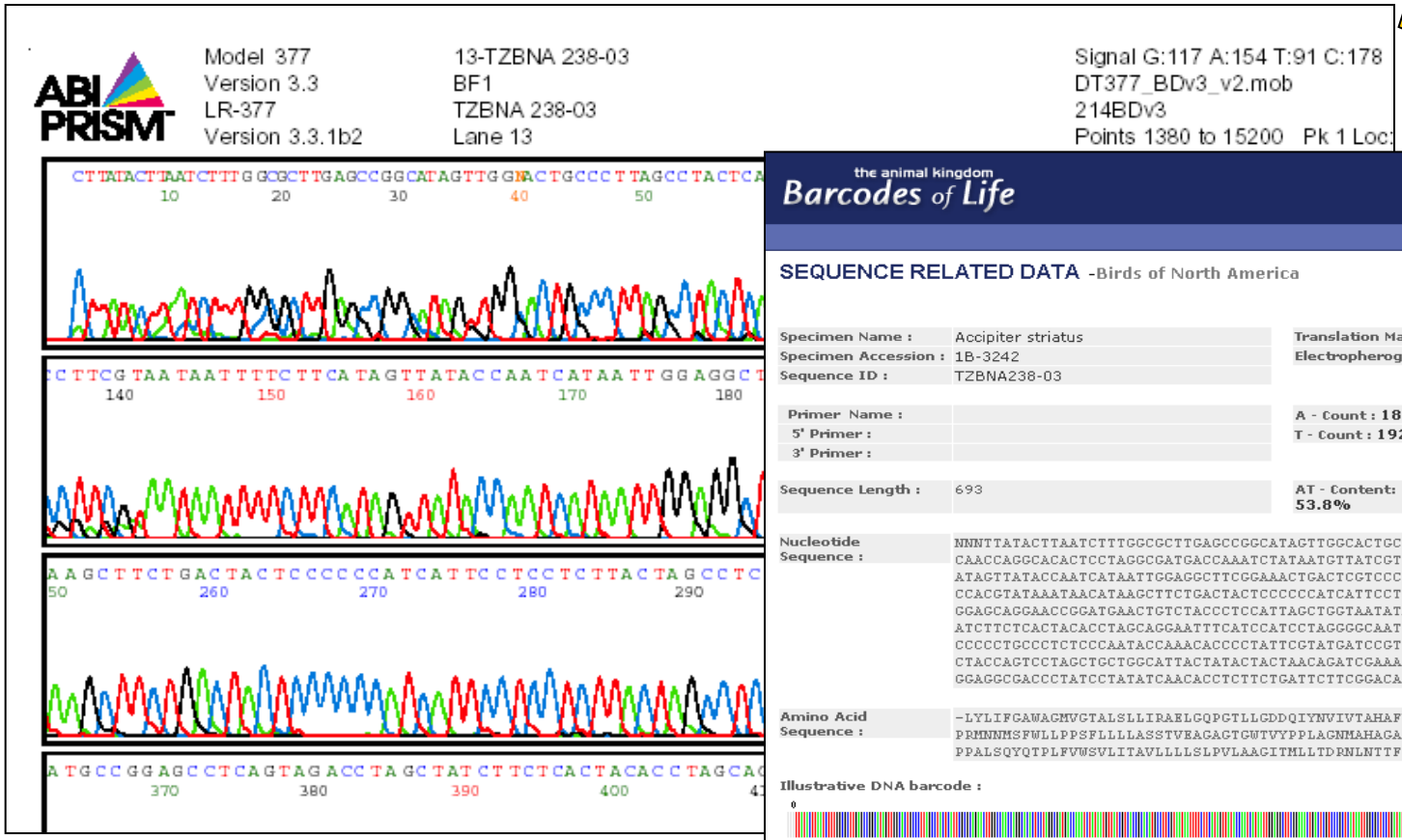
▲ **Collection Location [click on image to zoom]**

Dorsal View

Click on image to view images at larger size

©2003 Hebert Lab - Produced by CyberNatural Software

Sequence Data



the animal kingdom
Barcodes of Life

:: Print
:: Help

SEQUENCE RELATED DATA -Birds of North America

Specimen Name :	Accipiter striatus	Translation Matrix :	Vertebrate Mitochondrial Code
Specimen Accession :	1B-3242	Electropherogram :	Forward (BF1)
Sequence ID :	TZBNA238-03		
Primer Name :		A - Count : 181	G - Count : 105
5' Primer :		T - Count : 192	C - Count : 215
3' Primer :			
Sequence Length :	693	AT - Content: 53.8%	GC - Content : 46.2%

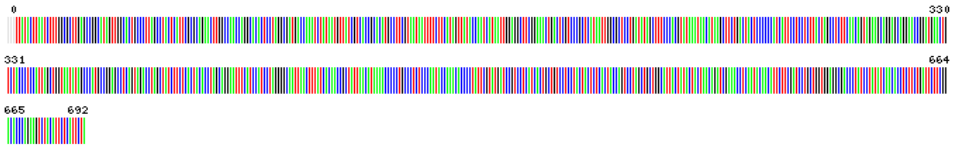
Nucleotide Sequence :

```
NNNTTATACTTAATCTTTGGCGCTTGAGCCGGCATAGTTGGCACTGCCCTTAGCCTACTCATTCGCGCAGAACTTGGC
CAACCAGGCACACTCCTAGGCGATGACCAAATCTATAATGTTATCGTCACCGCACATGCCTTCGTAATAATTTTCTTC
ATAGTTATACCAATCATAATTGGAGGCTTCGGAAACTGACTCGTCCCGCTCATAATTGGCGCTCCTGATATAGCTTTC
CCACGTATAAATAACATAAGCTTCTGACTACTCCCCCCATCATTCCTCCTCTTACTAGCCTCTTCAACAGTAGAAGCA
GGAGCAGGAACCGGATGAACTGTCTACCCTCCATTAGCTGGTAATATAGCCCATGCCGGAGCCTCAGTAGACCTAGCT
ATCTTCTCACTACACCTAGCAGGAATTTCATCCATCCTAGGGGCAATTAACTTTATCACAACCGCTATTAACATAAAA
CCCCCTGCCCTCTCCCAATACCAAACACCCCTATTCGTATGATCCGTCCTCATCACTGCTGTCCTCCTACTACTCTCA
CTACCAGTCCTAGCTGCTGGCATTACTATACTACTAACAGATCGAAACCTCAATACAACATTCTTCGATCCTGCTGGT
GGAGGCGACCCTATCCTATATCAACACCTCTTCTGATTCTTCGGACACCCAGAAGTCTACATTCTCATTCTA
```

Amino Acid Sequence :

```
-LYLIFGAWAGMVGTALSLLIRAELGQPGTLLGDDQIYNVIVTAHAFVMIFFMVMPIMIGGFGNWLVPLMIGAPDMAF
PRMNNMSFWLLPPSFLLLLASSTVEAGAGTGWTVYPPLAGNMAHAGASVDLAIFSLHLAGISSILGAINFITTAINMK
PPALSQYQTPLFVWSVLITAVLLLLSLPVLAAGITMLLTDRNLNTTFFDPAGGDPILYQHLFWFFGHPEVYILIL
```

Illustrative DNA barcode :



©2003 Hebert Lab - Produced by CyberNatural Software

Reality of DNA Barcoding

erflies, Saturniidae and Sphingidae.

Four pairs of overlooked Area de Conservacion Guanacaste (ACG) species flushed out of the inventory by barcoding. Top to bottom: left, *Prepona* demodiceDHJ01 (caterpillars eat Chrysobalanaceae); right, *Prepona* demodiceDHJ02 (caterpillars eat Fabaceae); left, *Cocytius* luciferDHJ01 (caterpillars found in ACG dry forest); right, *Cocytius* luciferDHJ02 (caterpillars found in ACG rain forest); left, *Udranomia* kikkawaiDHJ01 (caterpillars found in ACG dry forest); right, *Udranomia* kikkawaiDHJ02 (caterpillars found in ACG rain forest); left, *Eacles* imperialisDHJ01 (caterpillars found in ACG rain forest); right, *Eacles* imperialisDHJ02 (caterpillars found in ACG dry forest).

D.H. Janzen, et Al. (2009) Integration of DNA Barcoding into an inventory of complex tropical biodiversity. Molecular Ecology Resources 9 (Suppl.1), 1-26.

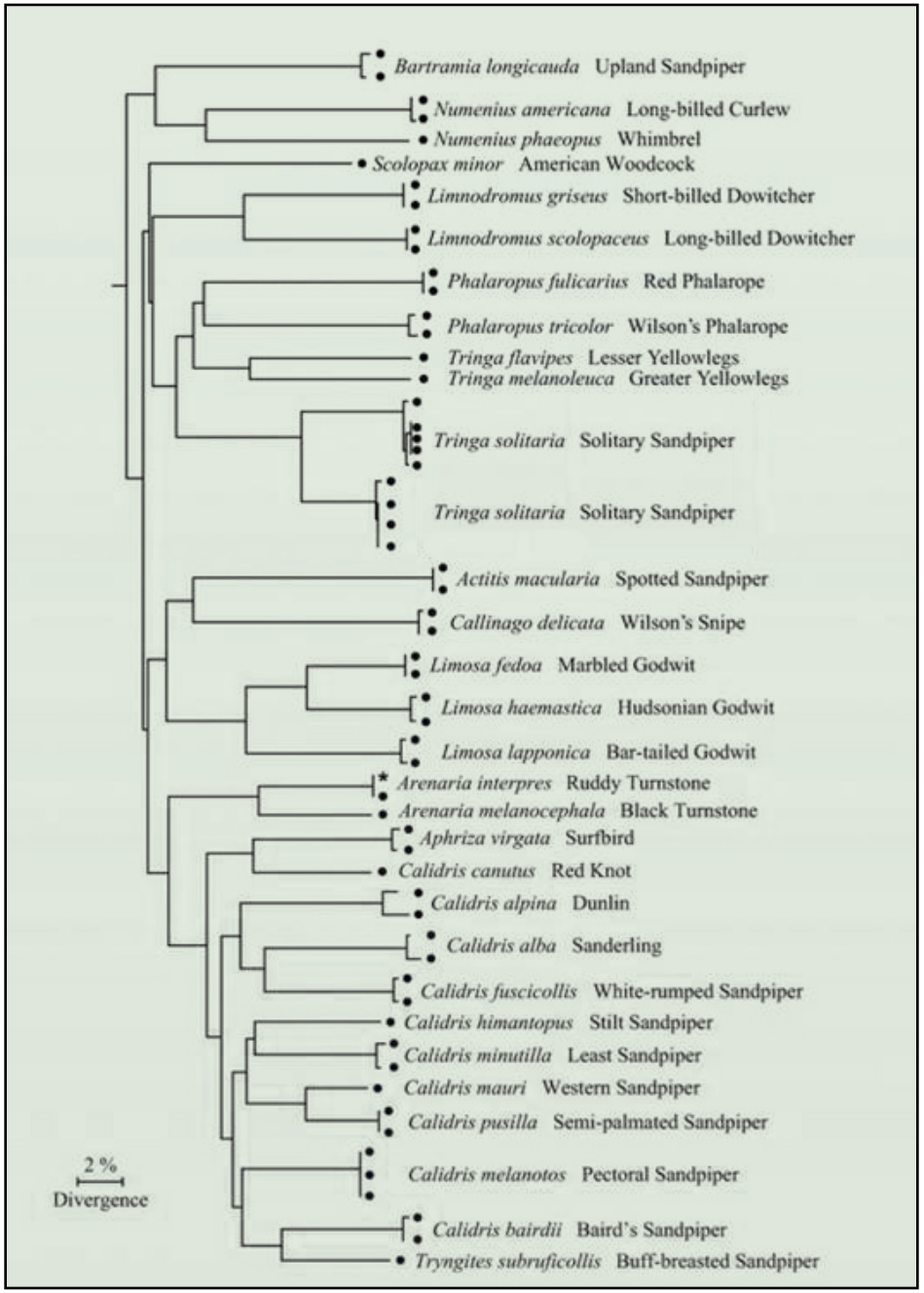
Fish Identification through DNA barcodes

Australian Fishes

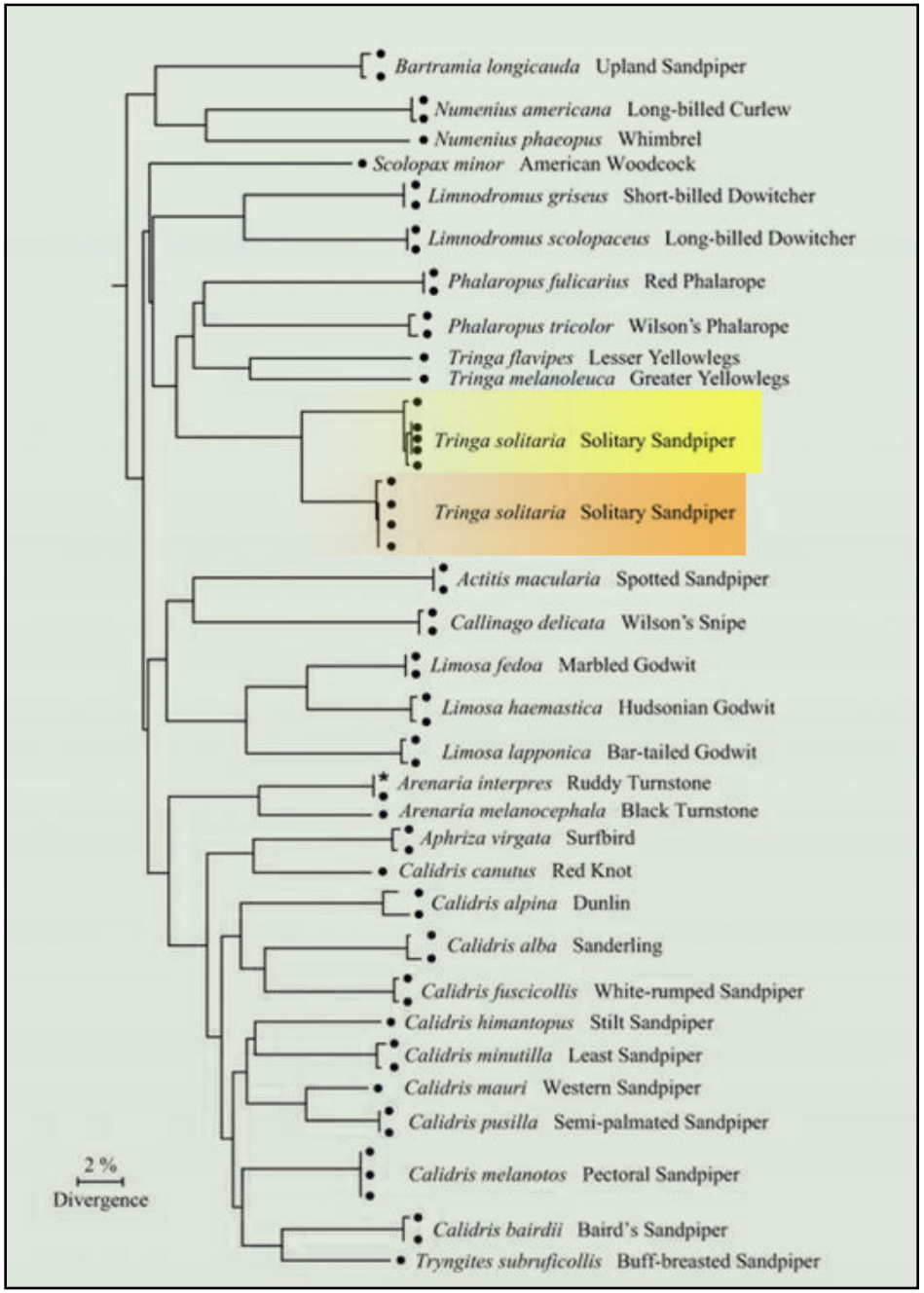
100% Resolution

.35% within species

9.93% between species



Bird Identification through DNA barcodes



Bird Identification through DNA barcodes

Now...who am I?

When DNA is combined with a comprehensive understanding of biology, the picture becomes clearer.

DNA

+

Morphology

+

Biology

Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*

Paul D. N. Hebert*†, Erin H. Penton*, John M. Burns‡, Daniel H. Janzen§, and Winnie Hallwachs§

*Department of Zoology, University of Guelph, Guelph, ON, Canada N1G 2W1; ‡Department of Entomology, National Museum of Natural History, Smithsonian Institution, Washington, DC 20560-0127; and §Department of Biology, University of Pennsylvania, Philadelphia, PA 19104

Contributed by Daniel H. Janzen, August 20, 2004

Astraptes fulgerator, first described in 1775, is a common and widely distributed neotropical skipper butterfly (Lepidoptera: Hesperiidae). We combine 25 years of natural history observations in northwestern Costa Rica with morphological study and DNA barcoding of museum specimens to show that *A. fulgerator* is a complex of at least 10 species in this region. Largely sympatric, these taxa have mostly different caterpillar food plants, mostly distinctive caterpillars, and somewhat different ecosystem preferences but only subtly differing adults with no genitalic divergence. Our results add to the evidence that cryptic species are prevalent in tropical regions, a critical issue in efforts to document global species richness. They also illustrate the value of DNA barcoding, especially when coupled with traditional taxonomic tools, in disclosing hidden diversity.

mitochondrial gene cytochrome *c* oxidase I (COI), whereas divergences among conspecific individuals average only 0.25% (11). Similar values were obtained in birds, with intraspecific divergences at COI averaging 0.27%, whereas congener divergences averaged 7.93% (14).

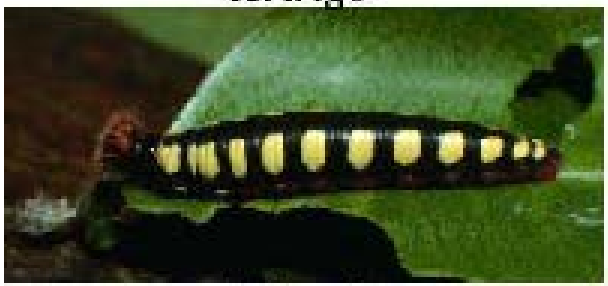
In this study, the addition of DNA barcodes to data on food plants, ecological distributions, caterpillar color patterns, and adult facies indicates that *A. fulgerator* consists of 10 largely sympatric species in the ACG. This result raises the prospect that, over its huge neotropical range, *A. fulgerator* may comprise many more hidden species. Imagine the biodiversity implications of this result for other wide-ranging, common, and "somewhat variable" species of neotropical animals and plants.

YESENN

SENNOV

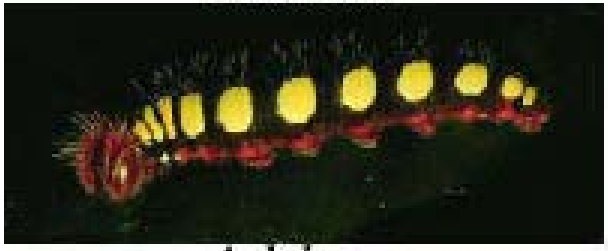
A. trigo

A. celt



A. loncho

A. ingcup

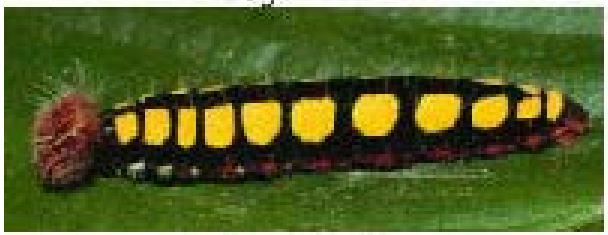


A. lohamp

A. lihamp

A. byttner

A. fabov



A. yesenn

A. sennov

...or even worse

When you return to the species, subtle morphology becomes evident

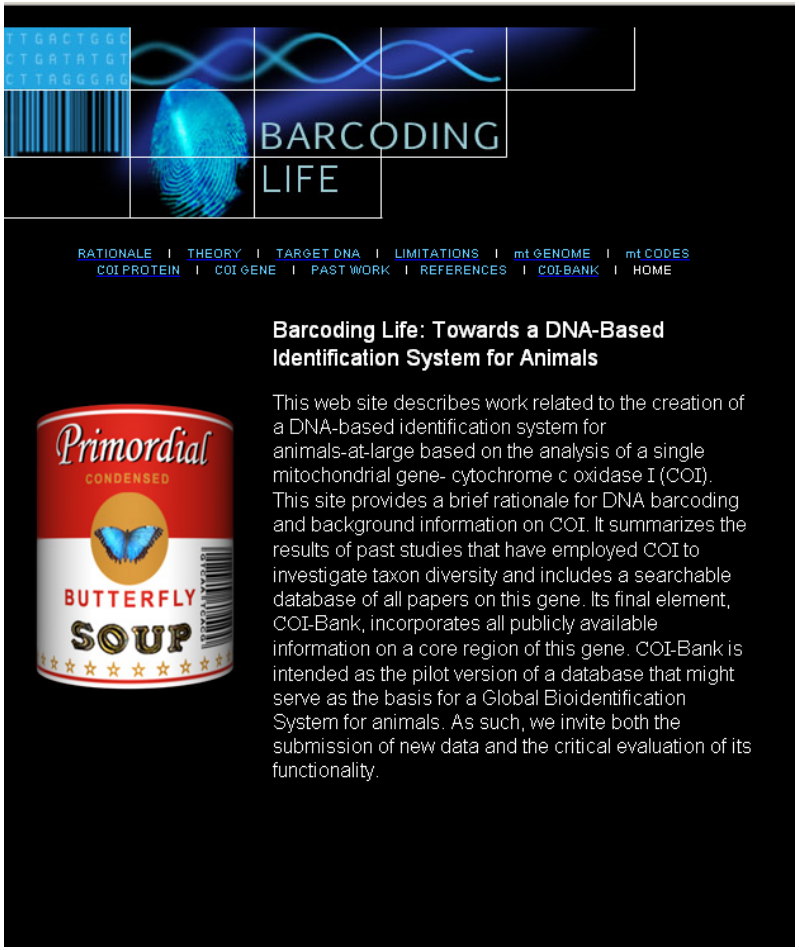
The Ultimate Dream

COI:
1 minute,
1 sequence,
1 name,
1 penny.
Radio Shack,
barcorder $7.99

D.H. Janzen

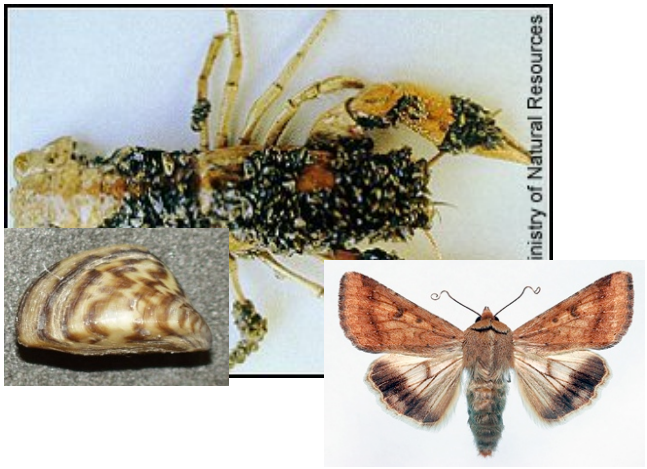
DNA Barcoding

* Is it just a new tool for taxonomy? Or might it also be a threat to traditional taxonomy?
* Why is its use generating controversy among practitioners, users, funders, etc.
* When does it work?
* When does it fail?
* Why and how can claims about its utility and accuracy be so different?



Why should we care?

Invasive Pests



Consumer Fraud

Human Health

Environmental Monitoring

Bird Strikes

Some Goals

* Build, over the next 20 years, a barcode library for all eukaryotic life
* Add alignment browser for visualing nucleotide and amino acid alignments
* Add capacity for multiple sequence markers
* Add nuclear markers
* Include plant markers: ITS, matK, and rbcL markers
* Develop algorithms for discovering new leaves on the “Tree of Life”
* Consider character-based identifications
* Develop probabilities and certainty values
* Develop algorithms for dealing with hybridization and lineage sorting (coalescence issues)

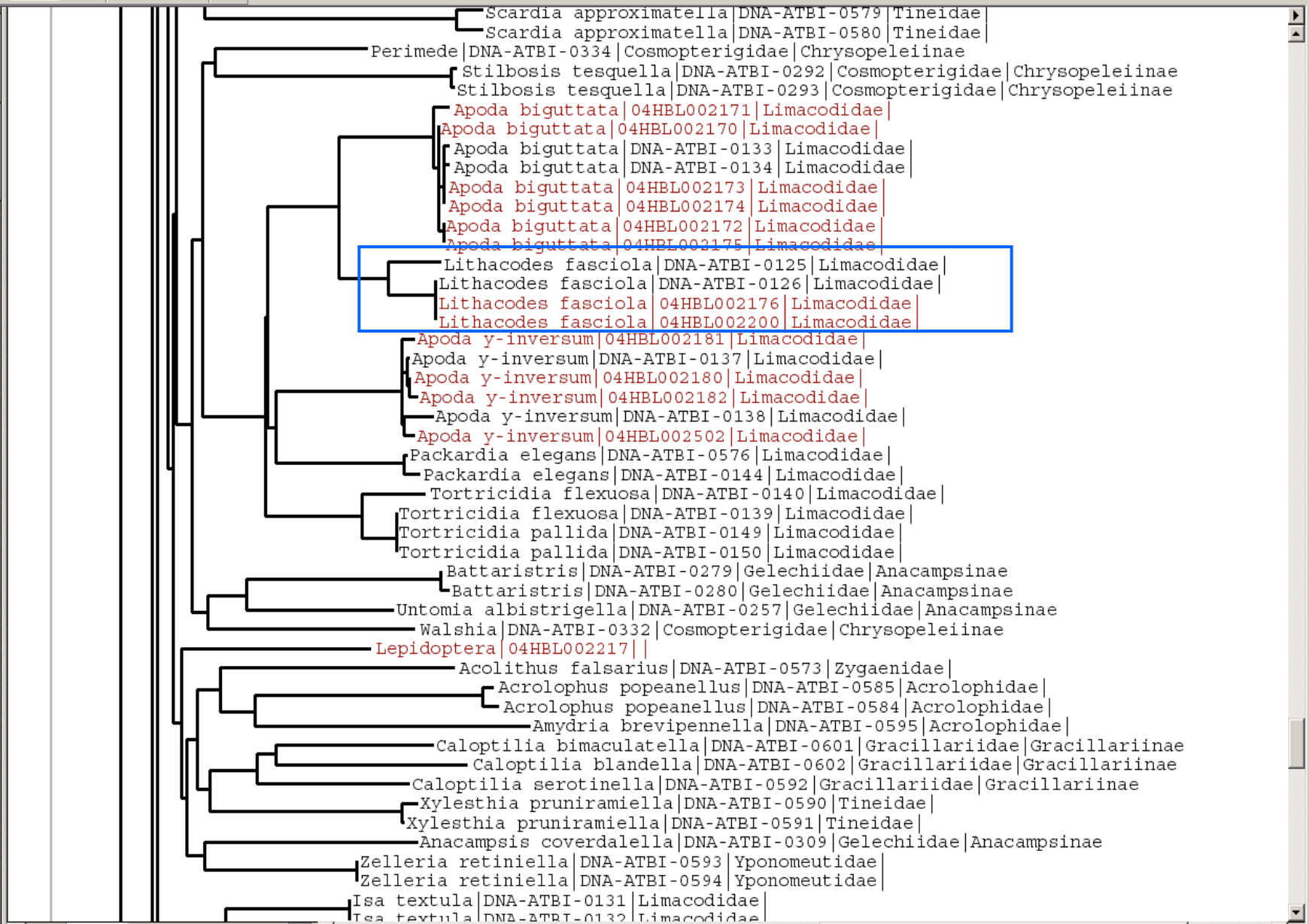
Acrobat Reader - [Smoky Mtns Lepidoptera -1198 sequences 428 species.pdf]

File Edit Document Tools View Window Help

250%

Bookmarks

Thumbnails



11 of 12 8.5 x 11 in

BARCODING LIFE: TEN REASONS TO IDENTIFY SPECIES BY DNA

1. Works with fragments
2. Works with all life stages
3. Genders are the same
4. Unmasks look-alikes
5. Reduces ambiguity
6. Makes expertise go further
7. Democratizes access
8. Open way for electronic handheld guide
9. Sprouts new leaves on the tree of life
10. Demonstrates value of collections
11. Speeds writing the encyclopedia of life

http://www.barcoding.si.edu/PDF/TenReasonsBarcoding.pdf

BARCODING LIFE: MORE REASONS TO IDENTIFY SPECIES BY DNA

12. Phylogenetic signal → a research agenda for taxonomists
13. A higher benchmark for type series
14. Enables research on minute taxa (microlepidoptera, nematodes, algae)

Some Problems

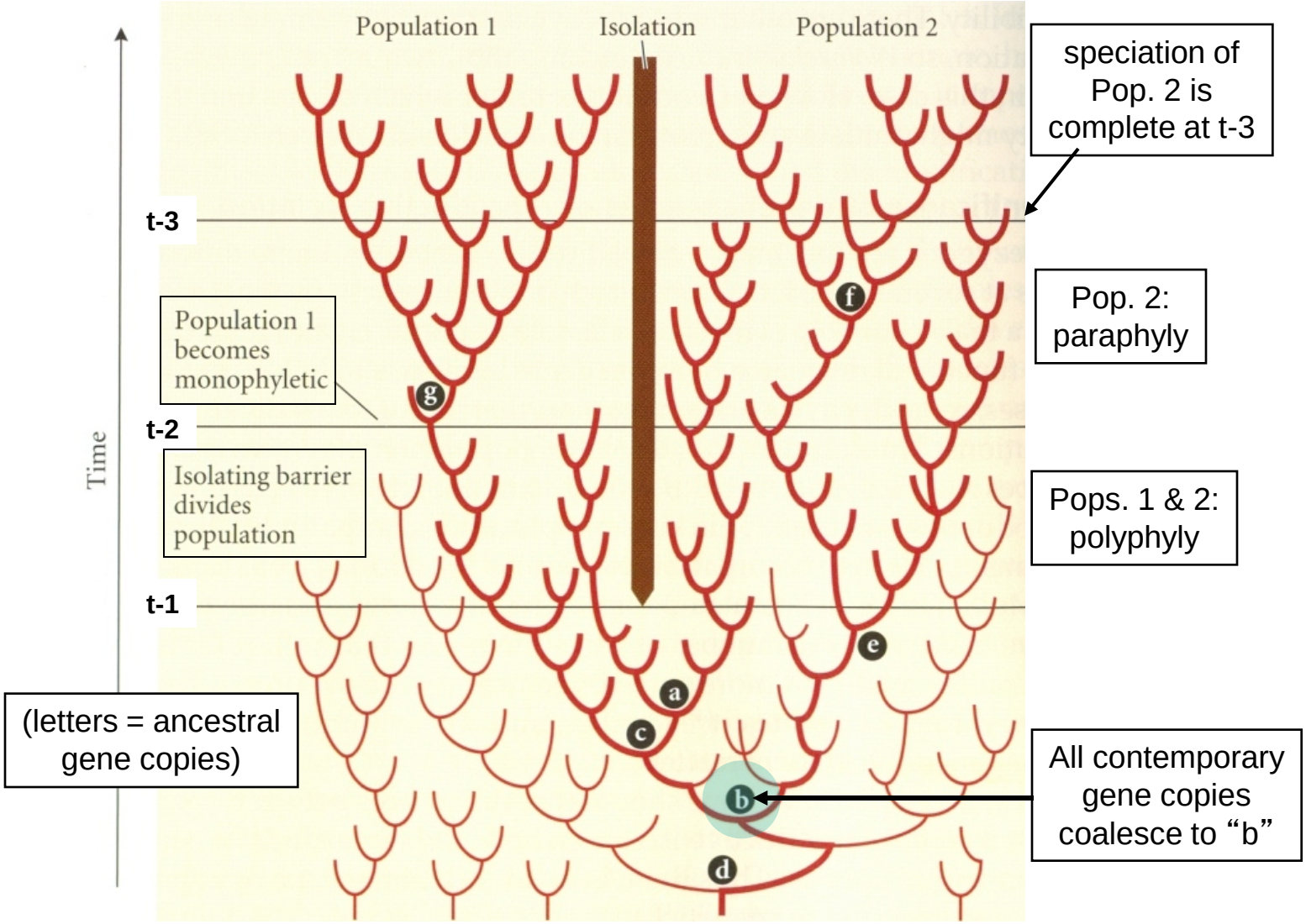
* ideally daughter species should be reciprocally monophyletic
- Funk and Omland (2003) ARES. 34: 397-423: only 70% of species in their examples were reciprocally monophyletic.
* **hybridization** is especially problematic: e.g., one haplotype is shared by 5 species of New England damselflies
  - important shortfall for many large genera
  - sloppy sex matters…many animals with external fertilization have evidence of hybridization

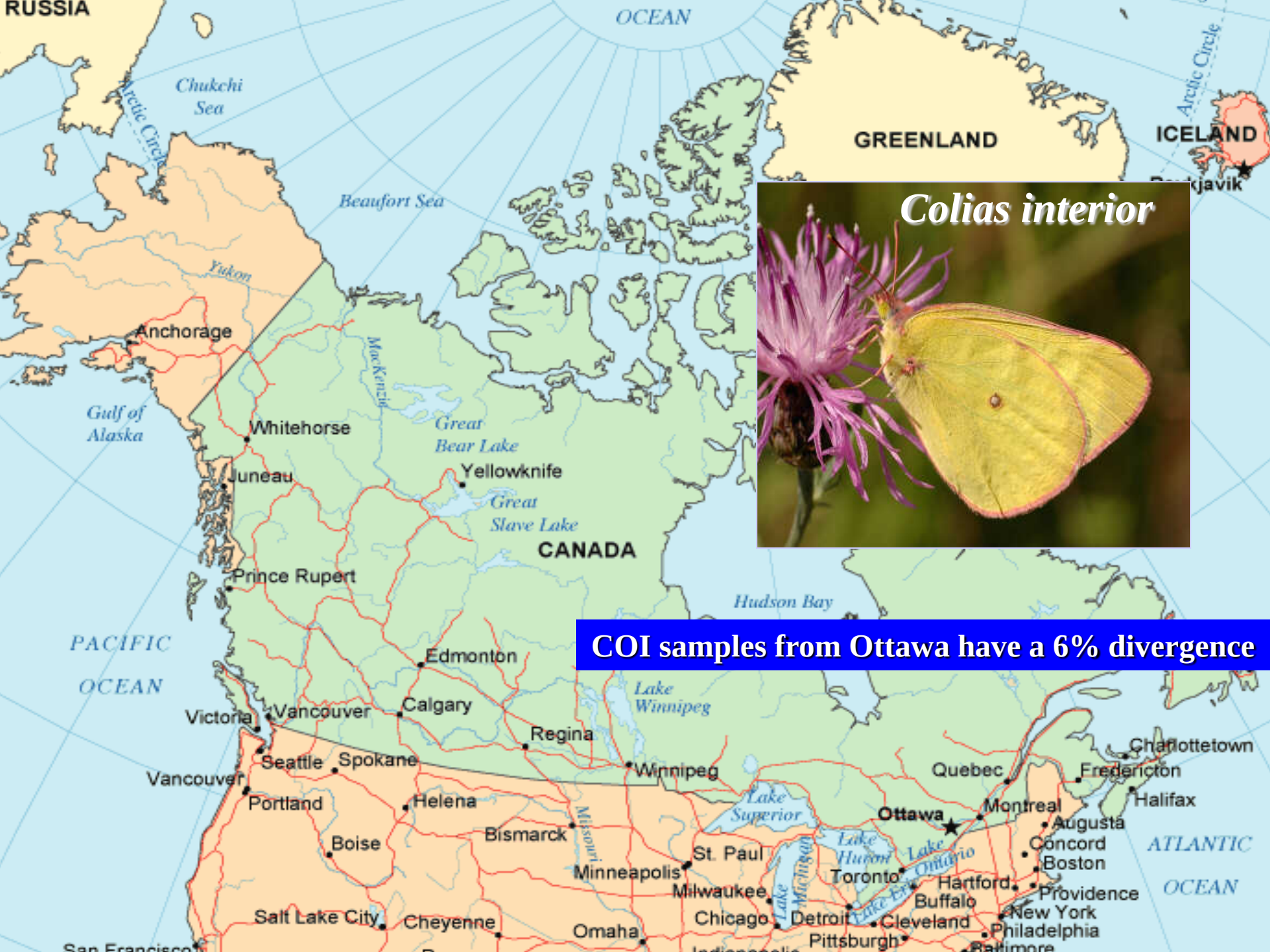
Some Problems

* ancestral polymorphism may be the most pervasive problem among closely related taxa
* an issue with recently speciating a taxa

* deep within-species splits possible: because haplotypes do not recombine, any biogeographic splits will linger (until haplotype extinction)
  e.g., sky-island populations from southern Appalachians usually different from Canadian haplotypes

Lineages of haplotypes at a single locus: transition from polyphyly to paraphyly to monophyly during speciation (Avise & Ball 1990)





Colias interior

COI samples from Ottawa have a 6% divergence

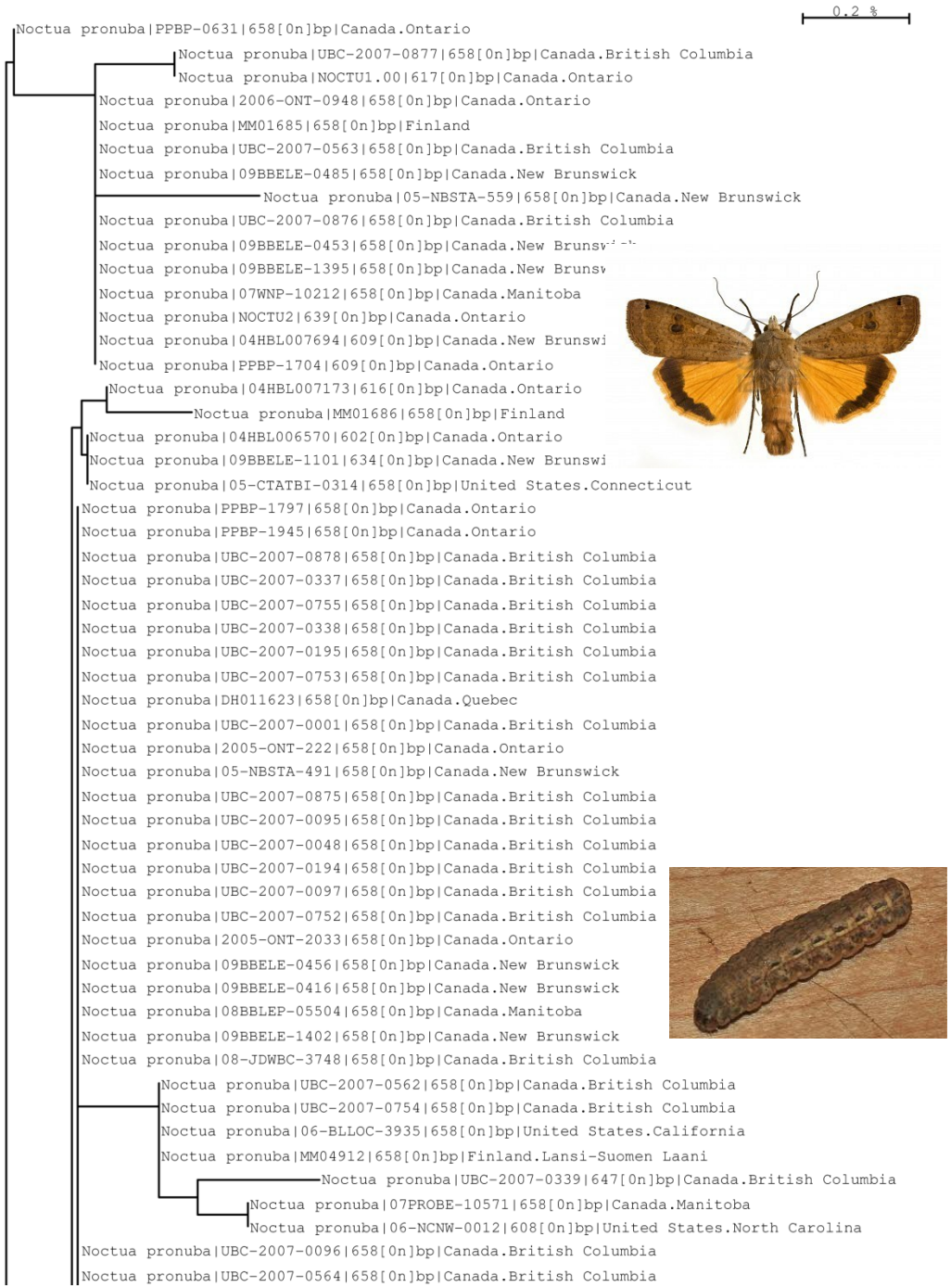
More Problems

* do barcodes work for asexual taxa?
* distances between taxa mean what?
* sometimes too little resolution (recent splits/taxa: e.g., lacewing song morphs) may not show up in COI
* sometimes too many haplotypes to make much sense (*Noctua pronuba*)

Mitochondrial Haplotypes in *Noctua pronuba*

My thought with the barcodes that that recently introduced species would most likely have only a single barcode, or maybe a few if multiple individuals were introduced, and maybe these would even tie them to a source like China, England, etc. We're not sure with some species, like *Scoliopteryx libatrix*, *Lycaena phlaeas*, *Pyrhhia exprimens*, whether they are naturally holoarctic or are introductions by man. So, I thought a really good test of this premise would be *Noctua pronuba* because we know the year (about 1979) when a single individual was found at Halifax, a port where many European insects have arrived by ship. It's systematic spread through the maritime provinces, then west and south in 200 to 300 mile increments until reaching Alaska, Florida, and Mexico in about 25 years. However, when I do the barcodes, I see there are - so far - 14 different haplotypes for it, several as divergent as 0.5 %. We are told that a divergence of 2 % usually is indicative of full species and takes likely millions of years. No one has adequately explained to me how a species with apparently a single introduction can have so many haplotypes, unless hundreds of individuals flew off the boat carrying many haplotypes.

Don Lafontaine Nov 2011



Noctua pronuba

BOLD website: www.Boldsystems.org
Consortium of the Barcode of Life: www.barcoding.si.edu
FISH-BOL: www.fishbol.org
Tephritid Fruitfly Initiative: http://www.barcoding.si.edu/tbi.html

Relevant Resources on Access and Benefits Sharing

CBD website (Bonn Guidelines and the CBD, GTI, and ABS National Focal Points): www.cbd.int/
Swiss Academy of Sciences: Access and Benefit Sharing - Good Practice for Academic Research: http://abs.scnat.ch

For more information please contact:

Consortium for the Barcode of Life Secretariat Office

CBOL is an international initiative hosted by the Smithsonian Institution and supported by the Alfred P. Sloan Foundation. As of early 2008, CBOL had over 170 Member Organization from more than 50 countries.

National Museum of Natural History
Smithsonian Institution
P.O. Box 37012 MRC 105
Washington, DC 20013-7012
CBOLInfo@si.edu
Tel: +1-202-633-0808; Fax: +1-202-633-2938
http://barcoding.si.edu

CONSORTIUM FOR THE BARCODE OF LIFE